

**ANALYSES OF THE DNA FLANKING THE HYPERVARIABLE  
MINISATELLITE MS31A (D7S21) IN THE MALAYSIAN POPULATION**

**YUSMIN BINTI MOHD. YUSUF**

A dissertation submitted to the Institute of Postgraduate Studies and Research,  
University of Malaya, in partial fulfilment of the requirements for the Degree of  
Master of Philosophy



**JUNE 1999**

**Institute of Postgraduate Studies and Research  
University of Malaya  
Kuala Lumpur**

Perpustakaan Universiti Malaya



A505623141

## Abstract

In this study, variations within the 5' flanking DNA region of the minisatellite MS31 (D7S21) were investigated. DNA samples of 310 unrelated individuals from the Malaysian population (103 Malays, 106 Chinese and 101 Indians) were obtained and assayed for base substitutional polymorphisms at positions -4, -220, and -108 relative to the MS31 minisatellite. The variations at positions -4, -220, and -108 resulted in the presence or absence of the *AluI*, *HgaI*, and *Psp1406I* restriction sites, respectively, and were detected by simple PCR-based RFLP assays. For each polymorphic site, the frequencies of the three possible genotypes (designated +/+, +/-, and -/-) in the three racial groups were calculated and tested for Hardy-Weinberg equilibrium distribution. In addition, the heterozygosity (**h**) and power of discrimination (**Pd**) values specific to each race were calculated.

Chi-square test results for all three sites did not show any significant deviation from those expected from Hardy-Weinberg equilibrium. The **h** values for the *AluI* site polymorphism ranged between 0.319 and 0.377, whilst **Pd** values ranged between 0.482 and 0.551. Similarly, the **h** values for the *HgaI* site polymorphism, ranged from 0.42 and 0.50, whilst the **Pd** values ranged between 0.56 and 0.63. Finally, for the *Psp1406I* polymorphisms, the **h** values ranged between 0.411 and 0.494, whilst the **Pd** values ranged between 0.565 and 0.620.

The chi square 2x2 contingency test was performed to determine if any association existed between different polymorphic sites (*HgaI-AluI* and *HgaI-Psp1406I*). Haplotype analysis was carried out to determine the phase of double

heterozygous individuals. Results from the contingency test indicated that there was significant association among the three polymorphic sites.

During the *Hga*I assay, seven samples of Malay and Indian origins produced irregular banding patterns that were not normally observed in the general population. The seven samples were sequenced. Analysis of the nucleotide sequences revealed a 12 bp deletion in all these samples, located from positions -230C to -241A.

## Abstrak

Dalam kajian ini, variasi yang terdapat pada kawasan 5' yang mengapit DNA bagi minisatelit MS31 (D7S21) diperiksa. Sampel DNA dari 310 individu yang tiada pertalian darah daripada populasi Malaysia (103 Melayu, 106 China dan 101 India) telah diperolehi dan dijalankan pemeriksaan untuk polimorfisma 'substitutional' bes pada kedudukan -4, -220, dan -108, iaitu kedudukan yang relatif kepada minisatellite MS31. Variasi yang terhasil pada posisi-posisi ini adalah disebabkan kehadiran atau ketiadaan tapak pemotongan bagi enzim *AluI*, *HgaI* dan *Psp1406I* melalui kaedah PCR yang ringkas. Bagi setiap tapak polimorphik, kekerapan bagi tiga kemungkinan genotype (dinamakan  $+/+$ ,  $+/-$  dan  $-/-$ ) dalam ketiga kumpulan kaum ini dikira dan diuji untuk keseimbangan taburan Hardy-Weinberg. Sebagai tambahan, heterozigositi ( $h$ ) dan kuasa diskriminasi ( $P_d$ ) juga dikira bagi setiap kaum.

Ujian chi-square menunjukkan ketiga-tiga tapak memberikan nilai yang bererti daripada nilai yang dijangkakan daripada keseimbangan Hardy-Weinberg. Nilai  $h$  bagi tapak polimorfisma *AluI* berada di antara 0.319 dan 0.377, manakala nilai  $P_d$  adalah di antara 0.482 dan 0.551. Nilai yang hampir serupa diperolehi untuk tapak polimorfisma *HgaI* iaitu nilai  $h$  di antara 0.42 dan 0.50, nilai  $P_d$  di antara 0.56 dan 0.63. Akhir sekali untuk tapak polimorphik *Psp1406I*, nilai  $h$  nya adalah di antara 0.411 dan 0.494, dan nilai  $P_d$  nya di antara 0.565 dan 0.602.

Ujian kontigensi 2x2 chi-square telah dijalankan untuk menentukan assosiasi yang wujud di antara tapak polimorphik yang berbeza (*HgaI-AluI* dan *HgaI-Psp1406I*). Analisa haplotip pula telah dijalankan untuk menentukan fasa bagi individu yang mempunyai heterozigus double. Keputusan yang di dapati menunjukkan assosiasi yang bererti di antara ketiga-tiga tapak polimorphik itu.

Semasa kajian *HgaI* dijalankan, di dapati tujuh sampel menunjukkan corak fragmen yang berbeza di-gel dibandingkan dengan kawalan positifnya. Sampel yang

berasal daripada kaum India dan Melayu ini diunjukkan. Analisa menunjukkan terdapat delesi sebanyak 12 bp pada semua sampel, yang berkedudukan -230C hingga ke -241A.

## **Acknowledgement**

I would like to express my deepest gratitude to my supervisors, Prof. Dr. Koh Chong Lek and Assoc. Prof. Dr. Sam Choon Kook for their invaluable advice, guidance and assistance throughout the course of my thesis. I am also greatly indebted to them for their time and patience in helping me.

I would also like to thank Miss Lim Moo Eng and Mr. Edmund for their assistance especially in collecting the blood sample.

To my ex-lab-mates and my present lab-mates, thank you for the support and friendship. I'll miss all of you.

Special thanks is dedicated to Professor Neva E Haites, and members of the DNA Lab, Medical Genetics, University of Aberdeen, for their advice and assistance during my stay in Aberdeen.

I would like to thank the University of Malaya for providing me the fellowship for my study, and the Ministry of Science, Technology and the Enviroment for the IRPA grant 03-07-04-115 to Prof. Koh Chong Lek, which supported this project.

Last but not least, special thanks is due to the members of my family especially my husband for the help, care, encouragement and understanding. Thank you all very much.

	Page
<b>Abstract</b>	i
<b>Abstrak</b>	iii
<b>Acknowledgement</b>	v
<b>Table of Contents</b>	vi
<b>Abbreviations</b>	xii
<b>List of Figures</b>	xvi
<b>List of Tables</b>	xviii

## **CHAPTER ONE: INTRODUCTION**

<b>1.1</b>	<b>: General introduction</b>	1
<b>1.2</b>	<b>: Genome organization</b>	1
1.2.1	: The human genome	1
1.2.2	: Repetitive DNA	4
1.2.3	: Hypervariable loci	6
<b>1.3</b>	<b>: Genetic polymorphisms and genetic studies</b>	7
1.3.1	: Conventional source of analysis	7
1.3.2	: Restriction fragment length polymorphism	8
1.3.3	: HVRs as an alternative typing system	10
<b>1.4</b>	<b>: Minisatellite: application to forensic studies</b>	11
1.4.1	: DNA fingerprinting	11
1.4.2	: DNA profiling	13
1.4.3	: Polymerase chain reaction (PCR) based approach	15
1.4.4	: Minisatellite repeat unit sequence variant	17
1.4.5	: Other PCR based DNA typing methods	21
1.4.6	: MS31	23
<b>1.5</b>	<b>: Objective of study</b>	28

## CHAPTER TWO: MATERIALS AND METHODS

<b>2.1</b>	<b>: Blood samples</b>	<b>30</b>
<b>2.2</b>	<b>: Materials</b>	<b>30</b>
<b>2.3</b>	<b>: Stock solutions and buffers</b>	<b>30</b>
2.3.1	: Stock solutions and buffers for DNA extraction	30
2.3.1.1	: 20x SSC	30
2.3.1.2	: 2x SSC	31
2.3.1.3	: 10% (v/w) SDS	31
2.3.1.4	: 80% (v/v) ethanol	31
2.3.1.5	: Buffered phenol-chloroform	31
2.3.1.6	: 2 M NaOAc, pH 7 or 5.6	32
2.3.2	: Stock solutions and buffers for 11.1x PCR mix	32
2.3.2.1	: 2 M Tris-base	32
2.3.2.2	: 2 M Tris-HCl, pH 7.5	32
2.3.2.3	: 2 M Tris-HCl, pH 8.8	32
2.3.2.4	: 1 M MgCl <sub>2</sub>	32
2.3.2.5	: 0.5 M EDTA	32
2.3.2.6	: 10 mM EDTA, pH 8.0	33
2.3.3	: Stock solutions and buffers for PCR	33
2.3.3.1	: 11.1x PCR mixture (mix)	33
2.3.3.2	: Primers	32



2.3.3.3	: <i>Taq</i> DNA polymerase	34
2.3.4	: Restriction endonucleases and buffers for digestion of PCR products	34
2.3.4.1	: Restriction enzymes and assay buffers	34
2.3.4.2	: 10% (w/v) Spermidine	35
2.3.5	: Stock solutions, buffers and gel for electrophoresis	35
2.3.5.1	: 5x TBE buffer	35
2.3.5.2	: EtBr (Ethidium bromide solution- 10 mg/ml)	36
2.3.5.3	: Agarose gel	36
2.3.5.4	: 6x Gel loading dye	36
2.3.5.5	: 30% (w/v) Acrylamide	37
2.3.5.6	: 10% (w/v) Ammonium persulfate	37
2.3.5.7	: 5% (w/v) Polyacrylamide gel	37
2.3.6	: Stock solutions and buffers for sequencing reaction	38
2.3.6.1	: 5 M Ammonium acetate	38
2.3.6.2	: 5x Sequenase buffers	38
2.3.6.3	: Termination mixes	38
2.3.6.4	: Chase mixes	38
2.3.6.5	: Labeling mixes	38
2.3.7	: Sterilisation	39
<b>2.4</b>	: <b>Extraction of DNA from blood cells</b>	39
<b>2.5</b>	: <b>O.D. 260/280 assay</b>	40

<b>2.6</b>	<b>: Flanking regions polymorphism assays</b>	<b>41</b>
2.6.1	: PCR	41
2.6.1.1	: <i>Alu</i> I +/- assay	42
2.6.1.2	: <i>Hga</i> I +/- assay	42
2.6.1.3	: <i>Psp</i> 1406I +/- assay	43
2.6.2	: Digestion of amplified products	43
2.6.3	: Electrophoresis	44
2.6.4	: Statistical analysis	44
<b>2.7</b>	<b>: Haplotype assays of flanking DNA polymorphisms</b>	<b>46</b>
2.7.1	: PCR	46
2.7.2	: Digestion of amplified products	47
2.7.3	: Electrophoresis	47
2.7.4	: Statistical analysis for haplotyping	48
<b>2.8</b>	<b>: Agarose gel electrophoresis of DNA</b>	<b>49</b>
<b>2.9</b>	<b>: Direct sequencing</b>	<b>49</b>
<b>2.10</b>	<b>: Flow charts</b>	<b>50</b>

## **CHAPTER THREE: RESULTS**

<b>3.1</b>	<b>: Flanking region polymorphisms assays</b>	<b>53</b>
3.1.1	: <i>Alu</i> I +/- assay	54
3.1.2	: <i>Hga</i> I +/- assay	58
3.1.3	: <i>Psp</i> 1406I +/- assay	62
<b>3.2</b>	<b>: Haplotype assays</b>	<b>66</b>

3.2.1	: <i>HgaI-AluI</i> assay	67
3.2.2	: <i>HgaI-Psp1406I</i> assay	67
<b>3.3</b>	<b>: Sequencing</b>	72
3.3.1	: Irregular heterozygous sample	72
3.3.2	: Irregular homozygous sample	72

## CHAPTER FOUR: DISCUSSION AND CONCLUSION

<b>4.1</b>	<b>: Introduction</b>	76
<b>4.2</b>	<b>: Minisatellite variant repeat mapping</b>	76
<b>4.3</b>	<b>: Flanking substitutional polymorphic sites analyses</b>	78
<b>4.4</b>	<b>: Haplotype assays</b>	84
<b>4.5</b>	<b>: Sequencing</b>	91
<b>4.6</b>	<b>: Conclusion</b>	95

<b>References</b>	96
-------------------	----

## Appendixes

Appendix A	: Pairwise comparisons by the heterogeneity G-test of the distributions of <i>AluI</i> + and -, <i>HgaI</i> + and -, and <i>Psp1406I</i> + and - alleles between different population samples	101
Appendix B	: Pairwise comparisons by the heterogeneity G-test of the distribution of four haplotypes between different population samples for <i>HgaI-AluI</i> and <i>HgaI-Psp1406I</i> assays	107
Appendix C	: Pairwise comparisons by the heterogeneity G-test of the distributions of eight haplotypes between different Malaysian population samples	111
Appendix D	: Results of MS31A (Locus D7S21) typing	113
Appendix E	: Results of <i>HgaI-AluI</i> assay	121



## Abbreviations

%	: percentage
(v/v)	: volume per volume
(w/v)	: weight per volume
<	: less than
>	: more than
A	: adenine
APS	: ammonium persulphate
ATP	: adenosine 5'-triphosphate
bp	: base pair
BPB	: bromophenol blue
BSA	: bovine serum albumin
C	: cytosine
cm	: centimetre
dATP	: deoxyadenosine 5'-triphosphate
dCTP	: deoxycytidine 5'-triphosphate
ddNTP	: dideoxyribonucleoside 5'-triphosphate
df	: degree of freedom
dGTP	: deoxyguanosine 5'-triphosphate
dH <sub>2</sub> O	: distilled water
DNA	: deoxyribonucleic acid
dNTP	: deoxyribonucleoside 5'-triphosphate
dsDNA	: double-stranded DNA

dTTP	: deoxythymidine 5'-triphosphate
EDTA	: ethylenediaminetetraacetate
EtBr	: ethidium bromide
EtOH	: ethanol
g	: gramme
G	: guanine
h	: heterozygosity
HCl	: hydrochloric acid
hr	: hour
HVR	: hypervariable region
kb	: kilobase pairs
M	: molar
mA	: milliamperere
Mb	: megabase pairs
mg	: milligramme
MgCl <sub>2</sub>	: magnesium chloride
min	: minute
ml	: millilitre
mm	: millimetre
mM	: millimolar
MVR-PCR	: minisatellite variant repeat mapping by the polymerase chain reaction
NaCl	: sodium chloride or natrium chloride

NaOAc	: sodium acetate or natrium acetate
NH <sub>4</sub> OAc	: ammonium acetate
OD	: optical density
p	: probability
PAGE	: polyacrylamide gel electrophoresis
Pd	: power of discrimination
psi	: pounds per square inch
RFLP	: restriction fragment length polymorphism
rpm	: revolutions per minute
sdH <sub>2</sub> O	: sterile-distilled water
SDS	: sodium dodecyl sulphate
sec	: second
SSC	: sodium chloride-sodium citrate
ssDNA	: single-stranded DNA
T	: thymine
TBE	: Tris-borate EDTA
Tris	: Tris (hydroxymethyl) methylamine
U	: unit
uv	: ultraviolet
VNTR	: variable number of tandem repeats
°C	: degree Celsius
µg	: microgramme
µl	: microlitre

$\mu\text{m}$  : micrometre

$\mu\text{M}$  : micromolar

$\chi^2$  : Chi square



## List Of Figures

Figure 1	: Organization of the human genome	2
Figure 2	: The polymerase chain reaction (PCR)	16
Figure 3	: The principle of minisatellite repeat coding (MVR-PCR)	20
Figure 4	: G-banding pattern of human chromosome 7	25
Figure 5	: Organization of the D7S21 locus	26
Figure 6	: Base sequences for locus D7S21 (MS31A at 5' end)	27
Figure 7	: Flowchart of work performed for flanking region assay and haplotyping of the minisatellite	51
Figure 8	: Steps involved in preparation of sample for sequencing	52
Figure 9	: Ethidium bromide-stained 4% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:1) of <i>AluI</i> +/- assay in MS31A 5' flanking DNA (-4A/G polymorphism)	55
Figure 10	: Ethidium bromide-stained 5% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:2) of <i>HgaI</i> +/- assay in MS31A 5' flanking DNA (-220 G/C polymorphism)	59
Figure 11	: Ethidium bromide-stained 4% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:1) of <i>Psp1406I</i> I +/- assay in MS31A 5' flanking DNA (-108C/T polymorphism)	63
Figure 12	: Ethidium bromide-stained 5% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:2) of <i>HgaI</i> - <i>AluI</i> haplotype assay	67
Figure 13	: Ethidium bromide-stained 5% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:2) of <i>HgaI</i> - <i>Psp1406I</i> haplotype assay	68
Figure 14	: Samples chosen from <i>HgaI</i> assay for sequencing	72
Figure 15	: This figure shows the position of the 12 bp deletion	73
Figure 16	: Comparison of base sequence between a normal heterozygous	

	sample and an irregular heterozygous sample	74
Figure 17	: Comparison of base sequence between a normal heterozygous sample and an irregular homozygous sample	75
Figure 18	: Regular and irregular heterozygous samples	93
Figure 19	: Regular and irregular homozygous samples	94

## List Of Tables:

Table 1	: Classes of tandemly repetitive DNA sequences	5
Table 2	: 11.1x PCR mix	33
Table 3	: Primers	34
Table 4	: Restriction endonucleases and buffers	34
Table 5	: Distribution and frequencies of <i>AluI</i> site genotypes and alleles in Malays, Chinese and Indians in Malaysian population	56
Table 6	: Distribution and frequencies of <i>HgaI</i> site genotypes and alleles in Malays, Chinese and Indians in Malaysian population	60
Table 7	: Distribution and frequencies of <i>Psp1406I</i> site genotypes and alleles in Malays, Chinese and Indians in Malaysian population	64
Table 8	: 2x2 contingency test for <i>HgaI-AluI</i> MS31A 5' flanking haplotype	69
Table 9	: 2x2 contingency test for <i>HgaI-Psp1406I</i> MS31A 5' flanking haplotype	70
Table 10	: Comparison of the allele frequencies at <i>AluI</i> and <i>HgaI</i> sites in the 5' flanking DNA of MS31A among the Malaysian, Caucasians and Japanese population samples	81
Table 11	: Pairwise comparisons by the heterogeneity G-test of the distributions of <i>AluI</i> + and – alleles and <i>HgaI</i> + and – alleles between different population samples	82
Table 12	: Comparison of the allele frequencies at the <i>Psp1406I</i> sites in the 5' flanking DNA of MS31A among the Malaysian population samples	83
Table 13	: Pairwise comparisons by the heterogeneity G-test of the distributions of <i>Psp1406I</i> + and – alleles between different population samples	83

Table 14	:Distribution of the four haplotype at the MS31A 5' <i>HgaI</i> and <i>AluI</i> sites in the Malay, Chinese, Indians, Caucasian and Japanese population samples	85
Table 15	: Pairwise comparisons by the heterogeneity G-test of the distributions of four haplotypes at the <i>HgaI</i> and <i>AluI</i> sites between different population samples	86
Table 16	: Distribution of the four haplotypes at the MS31A 5' flanking <i>HgaI</i> and <i>Psp1406I</i> sites in Malays, Chinese, and Indians from Malaysia	87
Table 17	: Pairwise comparisons by the heterogeneity G-test of the distributions of four haplotypes at <i>HgaI</i> and <i>Psp1406I</i> sites between different population samples	87
Table 18	: $\chi^2$ analysis of eight haplotypes at the MS31A flanking polymorphic sites in the Malaysian population	89
Table 19	: Distribution of the eight haplotypes at the MS31A 5' flanking polymorphic sites in Malaysian Malays, Chinese and Indians from Malaysia	90
Table 20	: Pairwise comparisons by the heterogeneity G-test of the distributions of eight haplotypes at the <i>HgaI</i> , <i>AluI</i> , and <i>Psp1406I</i> sites between different Malaysian population samples	90